



## مدل‌سازی پایداری خاکدانه‌ها با استفاده از ماشین‌های بردار پشتیبان و رگرسیون خطی چند متغیره

علی اصغر بسالت‌پور<sup>۱\*</sup> - حسین شیرانی<sup>۲</sup> - عیسی اسفندیارپور بروجنی<sup>۳</sup>

تاریخ دریافت: ۱۳۹۲/۰۳/۱۹

تاریخ پذیرش: ۱۳۹۴/۰۱/۱۶

### چکیده

ماشین‌های بردار پشتیبان، ابزاری امیدبخش برای روش‌های مبتنی بر آموزش می‌باشند که برای مسایل تشخیص الگو بر اساس کمینه کردن احتمال بروز اشتباه، ایجاد شده‌اند. در این پژوهش، کارایی روش ماشین‌های بردار پشتیبان در مدل‌سازی برآورد پایداری خاکدانه‌ها (از طریق محاسبه میانگین هندسی قطر خاکدانه‌ها، GMD) با رگرسیون خطی چند متغیره مقایسه شد. برای این منظور از داده‌های زودیافت مؤثر بر پایداری خاکدانه‌ها شامل برخی ویژگی‌های توپوگرافیکی، پوشش گیاهی و خاک استفاده گردید. برای بررسی کارایی مدل‌ها نیز از برخی شاخص‌های آماری نظیر ضریب همبستگی ( $r$ )، میانگین مربعات خطا (MSE) و درصد خطا (ERROR%) بین مقادیر اندازه‌گیری شده و برآورد شده استفاده شد. نتایج نشان داد که کارایی مدل‌های مبتنی بر ماشین‌های بردار پشتیبان در تخمین پایداری خاکدانه‌ها، بسیار بیشتر از روش رگرسیون خطی چند متغیره مرسوم بود. مقادیر MSE و  $r^2$  مدل ماشین‌های بردار پشتیبان طراحی شده برای برآورد پایداری خاکدانه‌ها به ترتیب، برابر ۰/۰۰۵ و ۰/۸۶ بودند. درصد خطا برای تخمین پایداری خاکدانه‌ها با استفاده از مدل ماشین‌های بردار پشتیبان برابر ۱۰/۷ درصد بود در حالی که مقدار درصد خطا برای مدل رگرسیونی برازش داده شده، حدود ۱۵/۷ درصد بود. بنابراین به نظر می‌رسد که بتوان از ماشین‌های بردار پشتیبان برای برآورد برخی ویژگی‌های فیزیکی خاک (نظیر میانگین هندسی قطر خاکدانه‌ها) در منطقه‌ی مورد بررسی استفاده نمود.

**واژه‌های کلیدی:** ساختمان خاک، ماشین‌های بردار پشتیبان، میانگین هندسی قطر خاکدانه‌ها، ویژگی‌های فیزیکی خاک

### مقدمه

خاکدانه‌ها<sup>۴</sup>، میانگین هندسی قطر خاکدانه‌ها<sup>۵</sup>، رس قابل پراکنش<sup>۶</sup> و مقاومت کششی خاکدانه‌ها<sup>۷</sup> اشاره نمود (۵ و ۹). با این وجود، بسیاری از این روش‌ها و اندازه‌گیری پایداری خاکدانه‌ها نسبتاً پیچیده و عمدتاً وقت‌گیر و پرهزینه می‌باشند و استفاده از آن‌ها در مقیاس‌های بزرگ (نظیر حوزه‌های آبخیز)، با محدودیت‌های فراوان مواجه است. به همین سبب، امروزه از روش‌های غیرمستقیم (مانند توابع انتقالی خاک<sup>۸</sup> و توابع تخمین مکانی خاک<sup>۹</sup>) به‌منظور برآورد این ویژگی‌ها استفاده می‌گردد. در این توابع با استفاده از داده‌های زودیافت مانند توزیع اندازه‌ی ذرات خاک، ماده‌ی آلی، کربنات کلسیم معادل، ویژگی‌های توپوگرافیکی (مانند درصد شیب، جهت شیب، ارتفاع) و

ساختمان خاک از جمله ویژگی‌های فیزیکی مهم خاک می‌باشد که از طریق اثر بر چرخه کربن آلی و سایر عناصر غذایی، دریافت، ذخیره و انتقال آب در خاک، مقاومت در برابر فرسایش، هدایت هیدرولیکی، تهویه، گرمای خاک و نیز از راه تأثیر بر مقاومت مکانیکی خاک و انتشار عناصر غذایی، بر رشد گیاه و تولید محصول اثر می‌گذارد. یکی از شاخص‌های رایج برای مطالعه و بیان وضعیت ساختمان خاک، پایداری خاکدانه‌ها می‌باشد (۱). روش‌های گوناگونی برای اندازه‌گیری پایداری خاکدانه‌ها وجود دارند. همچنین از شاخص‌های گوناگونی برای بیان پایداری خاکدانه‌ها استفاده می‌شود که از آن جمله می‌توان به شاخص میانگین وزنی قطر

- 4- Mean Weight Diameter, MWD
- 5- Geometric Mean Diameter, GMD
- 6- Water Dispersible Clay, WDC
- 7- Tensile strength
- 8- Pedotransfer Functions, PTFs
- 9- Soil Spatial Prediction Functions, SSPFs

۱، ۲ و ۳- به ترتیب استادیار، دانشیار و استادیار گروه علوم خاک، دانشکده کشاورزی، دانشگاه ولی عصر (عج) رفسنجان  
\*نویسنده مسئول: (Email: a.besalatpour@vru.ac.ir)

همکاران (۱۳) نیز امکان استفاده از SVMs برای تخمین هدایت هیدرولیکی خاک با استفاده از داده‌های زودیافت را مورد بررسی قرار دادند و دقت بالایی این روش برای برآورد هدایت هیدرولیکی خاک را گزارش کردند.

با این وجود، تاکنون پژوهش‌های اندکی با هدف استفاده از SVMs در علوم خاک (به ویژه برآورد پایداری خاکدانه‌ها) انجام شده‌است اگرچه با توجه به ماهیت ذاتی ویژگی‌های خاک، یعنی تغییرپذیری زمانی و مکانی زیاد و دیریافت بودن بسیاری از این ویژگی‌ها، به نظر می‌رسد که ماشین‌های بردار پشتیبان دارای توانایی بالایی برای مدل‌سازی این ویژگی‌ها باشند. در همین راستا، پژوهش حاضر با هدف اصلی بررسی امکان استفاده از این روش برای تخمین پایداری خاکدانه‌ها با استفاده از ویژگی‌های زودیافت خاک در یک حوزه آبخیز کوهستانی با توپوگرافی شدید و قابلیت دسترسی محدود که شدیداً در معرض خطر فرسایش قرار دارد (حوزه آبخیز بازفت، غرب استان چهارمحال و بختیاری) بررسی شد. همچنین قابلیت استفاده از ماشین‌های بردار پشتیبان در برآورد پایداری خاکدانه‌ها با روش مرسوم رگرسیون خطی چند متغیره<sup>۴</sup> مقایسه شد.

## مواد و روش‌ها

### منطقه‌ی مورد مطالعه

منطقه‌ی مورد بررسی، بخشی از حوزه آبخیز کارون شمالی واقع در غرب استان چهارمحال و بختیاری می‌باشد که در موقعیت جغرافیایی ۳۳° ۳۲' تا ۳۲° ۳۲' طول شرقی و ۳۱° ۳۹' تا ۳۳° ۳۹' عرض شمالی قرار گرفته است (شکل ۱). بخش اعظم منطقه مورد مطالعه (حدود ۵۰ درصد) کوهستانی بوده که مرتفع‌ترین نقطه آن، زردکوه با ارتفاع ۴۲۲۱ متر می‌باشد. ارتفاع متوسط منطقه، حدود ۲۲۰۰ متر است. شیب منطقه متغیر می‌باشد و بیشترین آن مربوط به کلاس شیب ۴۰ تا ۷۰ درصد است (معادل ۴۶ درصد از وسعت کل حوزه). حدود ۵۶ درصد از وسعت کل منطقه، دارای پوشش مرتعی (به‌طور عمده، دافنه و گون) و بقیه دارای پوشش جنگلی (به‌طور عمده، بلوط ایرانی) یا عاری از پوشش می‌باشد. دامنه تغییرات بارش و درجه حرارت سالیانه در منطقه، به ترتیب، ۵۰۰ تا ۱۴۰۰ میلی‌متر و ۸ تا ۲۰ درجه سلسیوس می‌باشند. منطقه مورد بررسی در زون شکسته‌ی زاگرس قرار دارد و قدیمی‌ترین رخنمون‌های منطقه مربوط به دوران پالئوزوئیک زیرین (پروکامبرین) تا دوره‌های فرسایشی و نهشتی کواترنر در قلمروی زاگرس می‌باشند. واحدهای فیزیوگرافی غالب در منطقه بر اساس مطالعات صحرائی و خصوصیات ظاهری خاک‌ها شامل کوه، تپه، فلات و تراس‌های بالایی، دشت دامنه‌ای، دشت آبرفتی و اراضی مخلوط و متفرقه می‌باشند (۴).

پوشش گیاهی (نظیر درصد پوشش گیاهی و شاخص‌های مرتبط با آن)، برخی خصوصیات دیریافت مانند ویژگی‌های فیزیکی و مکانیکی خاک تخمین زده می‌شوند (۳). از جمله روش‌های نوین برای ایجاد توابع انتقالی و توابع تخمین مکانی خاک، ماشین‌های بردار پشتیبان<sup>۱</sup> (SVMs) می‌باشند.

ماشین‌های بردار پشتیبان، جزء الگوریتم‌های طبقه‌بندی نظارت‌شونده<sup>۲</sup> هستند که قادر به پیش‌بینی تعلق یک نمونه به یک کلاس یا گروه هستند. این الگوریتم برای تفکیک دو کلاس از هم، از یک صفحه استفاده می‌کند به گونه‌ای که این صفحه از هر طرف بیشترین فاصله را با هر کلاس داشته باشد. نزدیک‌ترین نمونه‌های آموزشی به این صفحه، «بردارهای پشتیبان» نام دارند (۲۴، ۲۵ و ۲۶). در واقع، SVMs نخستین بار برای مسایل طبقه‌بندی پیشنهاد شدند اما با اصلاحاتی که در معادله‌ها و توابع آن‌ها انجام گرفت، امروزه برای برآوردهای رگرسیونی غیرخطی نیز کاربرد دارند. در واقع، ماشین‌های بردار پشتیبان ساخته شده با تابع کرنلی بر خلاف روش‌های رگرسیون خطی مرسوم که تنها قادر به شبیه‌سازی روابط خطی هستند، دارای قابلیت بالایی در تشخیص روابط غیرخطی و پیچیده‌ی بین متغیر(های) وابسته و متغیر هدف می‌باشند. همچنین این روش‌های مبتنی بر هوش مصنوعی می‌توانند الگوهای موجود بین پارامترهای ورودی و خروجی را بهتر استخراج نموده و تشخیص بهتری از روند تغییرات پارامترها داشته باشند چرا که توانایی قابل توجهی در استخراج پاسخ هدف از داده‌های پیچیده دارند (۴).

تاکنون از ماشین‌های بردار پشتیبان در پژوهش‌ها و زمینه‌های گوناگونی برای مدل‌سازی استفاده شده است. برای مثال لیو و همکاران (۱۶) قابلیت استفاده از شبکه‌های عصبی مصنوعی<sup>۳</sup> (ANNs)، رگرسیون چند متغیره گام به گام و SVMs برای برآورد ظرفیت تبادل کاتیونی خاک را مورد بررسی قرار دادند و بیان کردند که روش SVMs بسیار قوی‌تر از دیگر روش‌های مورد مطالعه بود. لامورسکی و همکاران (۱۴) ضمن بررسی قابلیت شبیه‌سازی تغییرات رطوبت خاک نسبت به زمان با استفاده از دو مدل HYDRUS-1D و SVMs گزارش کردند که مدل SVMs نسبت به مدل HYDRUS-1D دارای ضریب خطای کمتری بود. ونگ و همکاران (۲۷) برای شبیه‌سازی روان‌آب ماهیانه از SVMs استفاده نمودند و به قابلیت بالایی این روش در مقایسه با سایر روش‌های مورد بررسی خود اشاره کردند. توواراکاوی و همکاران (۲۲) ضمن مقایسه روش SVMs با ANNs برای برآورد پارامترهای هدایت هیدرولیکی خاک، دریافتند که ماشین‌های بردار پشتیبان دارای توانایی بالاتری در تعیین مقادیر بهینه پارامترهای هدایت هیدرولیکی خاک هستند. لامورسکی و

1- Support Vector Machines, SVMs

2- Supervised Learning Algorithms

3- Artificial Neural Networks; ANNs

4- Multiple Linear Regression, MLR



شکل ۱- موقعیت جغرافیایی منطقه مورد بررسی در حوزه‌ی آبخیز کارون شمالی  
Figure 1- Location of the study area in the Northern Karoun basin

(میانگین قطر سوراخ‌های الک بالایی و پایینی)،  $n$  تعداد الک‌ها و  $W_i$  نسبت وزن خاکدانه‌های روی هر الک به وزن کل خاک استفاده شده پس از کسر وزن ذرات شن و سنگ‌ریزه می‌باشد که به صورت زیر محاسبه می‌شود:

$$W_i = \frac{W_i - W_{i(s)}}{W_i - \sum_{l=1}^n W_{i(s)}} \quad (2)$$

که در آن،  $W_i$  وزن ذرات باقی‌مانده در دامنه  $i$ ،  $W_{i(s)}$  وزن ذرات شن و سنگ‌ریزه در دامنه  $i$  و  $W_i$  وزن خشک خاک می‌باشند.

#### تعیین ویژگی‌های توپوگرافیکی و پوشش گیاهی

برخی خصوصیات توپوگرافیکی نقاط نمونه‌برداری شده، نظیر درصد شیب، جهت شیب و ارتفاع با استفاده از نقشه مدل رقومی ارتفاع<sup>۳</sup> با اندازه ابعاد سلولی ۲۰ متر در ۲۰ متر در محیط نرم‌افزار ایلوپس<sup>۴</sup> تعیین شدند. شاخص نرمال‌شده تفاوت پوشش گیاهی<sup>۵</sup> نیز به عنوان معیاری از وضعیت پوشش گیاهی نقاط مورد مطالعه استفاده شد. این شاخص از تصویر ماهواره‌ای IRS ۲۰۰۸ با وضوح ۲۴ متر در ۲۴ متر و در محیط نرم‌افزار ایلوپس استخراج گردید.

#### تجزیه و تحلیل آماری داده‌ها

پس از اندازه‌گیری و تعیین ویژگی‌های خاک، توپوگرافی و پوشش

#### نمونه‌برداری و اندازه‌گیری ویژگی‌های خاک

منطقه مورد مطالعه ابتدا بر اساس ویژگی‌های توپوگرافیکی، زمین‌شناسی و پوشش گیاهی به واحدهای کاری همسان تقسیم شد. سپس، نمونه‌برداری خاک از عمق سطحی (صفر تا ۵ سانتی‌متر) به صورت تصادفی نظارت شده<sup>۱</sup> انجام شد و در مجموع، ۱۶۰ نمونه با در نظر گرفتن سطح هر واحد کاری برداشت گردید. نمونه‌های خاک به آرامی در ظروف پلاستیکی قرار گرفتند تا کمترین آسیب به خاکدانه‌ها وارد شود. نمونه‌های خاک پس از انتقال به آزمایشگاه هواخشک شدند و برای انجام آزمایش‌های فیزیکی و شیمیایی لازم از الک دو میلی‌متری عبور داده شدند. ماده آلی خاک به روش والکی و بلاک (۱۷)، کربنات کلسیم معادل خاک به روش تیتراسیون برگشتی (۱۸)، بافت خاک به روش پیپت (۱۰) و درصد ذرات شن ریز و شن خیلی ریز با عبور سوسپانسیون آب و خاک به ترتیب از الک‌های ۱۸۰ و ۲۷۰ میکس<sup>۲</sup> تعیین شدند. برای اندازه‌گیری پایداری خاکدانه‌ها نیز ابتدا خاک هواخشک شده (بدون کوبیدن) از الک ۸ میلی‌متری عبور داده شد و سپس از روش ون باول (۲۳) اصلاح‌شده توسط کمپر و روزنو (۱۱) استفاده گردید و میانگین هندسی قطر خاکدانه‌ها (GMD) به عنوان شاخصی از پایداری خاکدانه‌ها با استفاده از رابطه<sup>۱</sup> تعیین شد.

$$GMD = \exp\left(\sum_{i=1}^n w_i \log \bar{x}_i\right) \quad (1)$$

که در آن،  $\bar{x}_i$  میانگین قطر خاکدانه‌های باقی‌مانده بر روی هر الک  $i$

3- Digital Elevation Model, DEM

4- Integrated Land and Water Information System, ILWIS

5- Normalized Difference Vegetation Index, NDVI

1- Supervised random soil sampling

2- Mesh

به‌منظور مفهوم‌سازی چگونگی عملکرد ماشین‌های بردار پشتیبان، نقاط آموزشی به‌صورت رابطه ۵ در نظر گرفته شدند:

$$D = \{(x_i, y_i)\}_{i=1}^n \quad (5)$$

که در این رابطه،  $x_i$  بردار ورودی بوده که  $x_i \in X$  و  $y_i$  مقدار مطلوب می‌باشد که  $y_i \in R$  است و  $n$  تعداد کل الگوهای داده‌ای می‌باشد. اگر داده‌ها به‌صورت خطی قابل تفکیک باشند؛ آن‌گاه قواعد تصمیم‌گیری که توسط یک صفحه‌ی بهینه که طبقات تصمیم‌گیری باینری را تفکیک می‌کند به‌صورت معادله زیر خواهد بود:

$$y = \text{sign}\left(\sum_{i=1}^n y_i a_i (X \cdot X_i) + b\right) \quad (6)$$

که در آن،  $y$  خروجی معادله،  $y_i$  ارزش طبقه‌ی نمونه آموزشی، بردار  $x = (x_1, x_2, \dots, x_n)$  نشان‌دهنده‌ی یک داده‌ی ورودی و بردارهای  $X_i$  ( $i = 1, \dots, N$ )، بردارهای پشتیبان هستند. در این معادله، پارامترهای  $a_i$  و  $b$  تعیین‌کننده‌ی ابر صفحه هستند.

اگر داده‌ها به‌صورت خطی قابل تفکیک نباشند، معادله‌ی ۶ به معادله‌ی زیر تغییر می‌یابد:

$$y = \text{sign}\left(\sum_{i=1}^n y_i a_i K(X, X_i) + b\right) \quad (7)$$

که تابع  $K(X, X_i)$  تابع کرنلی<sup>۲</sup> است که برای ایجاد ماشین‌هایی با انواع مختلفی از سطوح تصمیم‌گیری غیرخطی در فضای داده‌ها، ضرب‌های داخلی تولید می‌کند (۲۴ و ۲۵). به‌عنوان مثال، سه نوع تابع کرنل که در مدل SVM به‌کار می‌روند، عبارتند از:

الف) ماشین چند جمله‌ای با تابع کرنل

$$K(X, X_i) = (X \cdot X_i + 1)^d \quad (8)$$

که در آن،  $d$  درجه‌ی کرنل چند جمله‌ای است.

ب) ماشین تابع پایه‌ی شعاعی با تابع کرنل

$$K(X, X_i) = \exp(-1/\mathcal{G}^2 (X - X_i)^2) \quad (9)$$

که در آن،  $\mathcal{G}$  پهنای باند کرنل تابع پایه‌ی شعاعی است.

پ) ماشین دو لایه با تابع کرنل

$$K(X, X_i) = S[(X \cdot X_i)] = 1/[1 + \exp\{v(X \cdot X_i) - c\}] \quad (10)$$

که در آن،  $c$  و  $v$  پارامترهای تابع زیگموییدی  $S[(X \cdot X_i)]$  هستند، به‌گونه‌ای که نامعادله‌ی  $c \geq v$  برقرار باشد (۲۶). شکل ۲ فرآیند مدل SVM را نشان می‌دهد. مدل‌سازی با ماشین‌های بردار پشتیبان در این پژوهش در محیط نرم‌افزار کلمنتاین<sup>۳</sup> انجام شد.

گیاهی نقاط مورد مطالعه، برخی خصوصیات آماری داده‌ها نظیر بیشینه، کمینه، میانگین، واریانس و انحراف معیار با استفاده از نرم افزار SPSS تعیین شدند. سپس به‌منظور یکسان نمودن ارزش داده‌ها پیش از استفاده در آزمون گاما و مدل‌سازی با شبکه‌های عصبی، داده‌ها با استفاده از رابطه‌ی ۳ استاندارد شدند:

$$X_n = \left[ \left( \frac{X_i - \bar{X}}{X_{\max} - X_{\min}} \right) \times 0.5 \right] + 0.5 \quad (3)$$

که در این رابطه،  $X_n$  داده‌های استاندارد شده،  $X_i$  داده‌های مشاهده‌ای،  $X_{\max}$  و  $X_{\min}$  به‌ترتیب، بیشینه و کمینه‌ی داده‌های مشاهده‌ای و  $\bar{X}$  میانگین داده‌های مشاهده‌ای می‌باشند. این رابطه، داده‌ها را در بازه‌ی [۰ و ۱] تبدیل می‌کند.

### مدل‌سازی با رگرسیون خطی چند متغیره (MLR)

رگرسیون‌های خطی از جمله روش‌های ساده‌ی مدل‌سازی می‌باشند که از آن‌ها می‌توان برای پیش‌بینی یک یا چند متغیر پاسخ (وابسته) از مجموعه‌ای از مقادیر متغیرهای پیش‌بینی کننده (مستقل) استفاده نمود. رگرسیون‌های خطی به‌صورت دو متغیره و چند متغیره می‌باشند که در رگرسیون دو متغیره، یک متغیر مستقل و یک متغیر وابسته وجود دارد. در رگرسیون چند متغیره نیز، یک متغیر وابسته و دو یا چند متغیر مستقل وجود دارد. معادله‌ی کلی یک رگرسیون خطی چند متغیره به‌صورت رابطه ۴ می‌باشد:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_n X_n \quad (4)$$

که در آن،  $Y$  متغیر وابسته،  $X$  متغیر مستقل و  $\beta_1, \dots, \beta_n$  ضرایب رگرسیونی می‌باشند. مدل‌سازی رگرسیونی در محیط نرم افزار SPSS انجام شد.

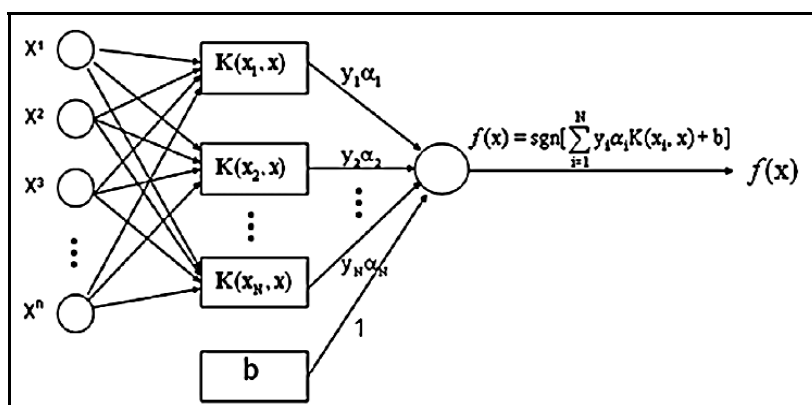
### مدل‌سازی با ماشین‌های بردار پشتیبان (SVMs)

ماشین‌های بردار پشتیبان (SVMs) از جمله روش‌های یادگیری نظارت‌شونده<sup>۱</sup> می‌باشند که از آن‌ها برای طبقه‌بندی و رگرسیون استفاده می‌شود (۲۴ و ۲۵). الگوریتم SVM نوع خاصی از مدل‌های خطی را می‌یابد که حاشیه‌ی ابر صفحه بیشینه شود. بیشینه کردن حاشیه‌ی ابر صفحه، منجر به حداکثر شدن تفکیک بین طبقات می‌شود. در واقع، نزدیک‌ترین نقاط آموزشی به حداکثر حاشیه‌ی ابر صفحه را بردارهای پشتیبان می‌نامند که از آن‌ها برای مشخص کردن مرز بین طبقات استفاده می‌شود. اگر داده‌ها به‌صورت خطی قابل جداسازی باشند، از ماشین‌های خطی برای تولید یک سطح بهینه که داده‌ها را بدون خطا و با حداکثر فاصله‌ی میان صفحه و نزدیک‌ترین نقاط آموزشی (بردارهای پشتیبان) تفکیک می‌نماید، استفاده می‌شود.

2- Kernel function

3- Clementine

1- Supervised learning



شکل ۲- فرآیند مدل SVM (بردار ورودی،  $b$  بایاس،  $[y_1\alpha_1, y_2\alpha_2, \dots, y_n\alpha_n]$  بردار وزن ها و  $f(x)$  معادله تصمیم گیری)  
 Figure 2- The SVM model schematic ( $[X=[x^1, x^2, \dots, x^n]$ : input vector;  $b$ : bias;  $[y_1\alpha_1, y_2\alpha_2, \dots, y_n\alpha_n]$ : weights vector; and  $f(x)$ : decision equation)

### شاخص‌های ارزیابی مدل‌ها

به منظور بررسی کارایی و دقت دو مدل طراحی شده برای برآورد پایداری خاکدانه‌ها در این پژوهش، از برخی شاخص‌های آماری نظیر ضریب همبستگی ( $r$ )، میانگین مربعات خطا<sup>۱</sup> (MSE) و درصد خطا (ERROR%) استفاده شد که معادله‌ی دو شاخص آخر به صورت زیر می‌باشند:

$$MSE = \frac{1}{n} \sum_{i=1}^n [y_k - \hat{y}_k]^2 \quad (11)$$

$$ERROR\% = \frac{\sum_{k=1}^n |y_k - \hat{y}_k|}{\sum_{k=1}^n y_k} \times 100 \quad (12)$$

که در این روابط،  $y_k$  و  $\hat{y}_k$  به ترتیب، مقادیر مشاهده‌ای و برآورد شده و  $n$  تعداد کل مشاهدات می‌باشند.

### نتایج و بحث

#### ویژگی‌های خاک، توپوگرافی و پوشش گیاهی

جدول ۱ توصیف آماری ویژگی‌های خاک، توپوگرافی و پوشش گیاهی نقاط مورد مطالعه استفاده شده برای برآورد پایداری خاکدانه‌ها را نشان می‌دهد. بر اساس توزیع اندازه‌ی ذرات خاک، بافت غالب خاک در منطقه شامل لوم رس سیلتی، لوم سیلتی، لوم و رس سیلتی بود. ماده‌ی آلی خاک بین ۰/۲۲ تا ۶/۳۳ درصد و کربنات کلسیم معادل خاک، به ترتیب دارای مقادیر کمیته، بیشینه، میانگین و انحراف معیار ۱/۰۰، ۸۰/۶۵، ۲۵/۳۲ و ۲۱/۱۲ درصد بود. مقادیر زیاد کربنات

کلسیم معادل خاک در منطقه مورد مطالعه به دلیل وجود مواد مادری مارن غنی از گچ و آهک و وجود افق کلسیک می‌باشد (۱۷). شاخص NDVI در نقاط نمونه‌برداری از ۰/۳۱- تا ۰/۴۷ متغیر بود و دارای میانگین و انحراف معیار به ترتیب، ۰/۱۳ و ۰/۰۸ بود. شاخص NDVI بیان‌کننده‌ی سبزینه گیاهی است که در ارتباط با درصد تابش‌های جذب شده فتوسنتزی می‌باشد و نمایان‌گر فعالیت کلروفیل در گیاهان است. این شاخص بین ۱- تا ۱ متغیر بوده که افزایش میزان عددی آن، بیان‌گر افزایش درصد پوشش گیاهی سبز سالم می‌باشد و بنابراین سطوح بدون پوشش گیاهی، دارای مقادیر منفی NDVI هستند (۱۲). درصد شیب در نقاط نمونه‌برداری بین ۳/۱۰ تا ۸۰/۶۱ درصد با میانگین ۲۸/۳۱ درصد متفاوت بود که نمایان‌گر وجود توپوگرافی شدید در منطقه‌ی مورد مطالعه است. جهت شیب، دارای میانگین ۱۷۱/۸۲ رادیان و ارتفاع نقاط نمونه‌برداری به ترتیب دارای مقادیر کمیته، بیشینه و میانگین ۹۰۷/۰۰، ۲۶۳۹/۰۰ و ۱۷۱۴/۵۳ متر بود.

در بسیاری از مطالعات، برخی از ویژگی‌های خاک (مانند توزیع اندازه‌ی ذرات خاک، ماده‌ی آلی و کربنات کلسیم معادل) و نیز برخی از خصوصیات توپوگرافی و پوشش گیاهی (مانند درصد و جهت شیب، ارتفاع و درصد پوشش گیاهی) به عنوان پارامترهای اثرگذار بر پایداری خاکدانه‌ها نام برده شده‌اند (۱، ۲، ۶، ۷، ۱۹ و ۲۱). برای مثال، سیکس و همکاران (۱۹) و آندسودان و همکاران (۲) بیان کردند که کمبود و کاهش ماده‌ی آلی سبب کاهش اندازه‌ی خاکدانه‌ها، تخریب آن‌ها و در نتیجه، حساسیت خاک به فرسایش می‌شود. در واقع، ماده‌ی آلی از دو راه افزایش ویژگی آب‌گریزی و افزایش چسبندگی بین ذرات خاک، سبب افزایش مقاومت خاکدانه‌ها در برابر خیس شدن و پایداری ساختمان خاک می‌گردد (۸). تیسدال و اودز (۲۱) رس، اکسیدهای آهن و آلومینیوم و کربنات‌ها را بر پایداری ساختمان خاک مؤثر دانستند.

1 -Mean Square Error, MSE

جدول ۱- خلاصه آماری پارامترهای مورد استفاده برای مدل‌سازی پایداری خاکدانه‌ها

Table 1- Description statistics of the parameters used for modelling of aggregate stability

پارامتر Parameter	شاخص آماری Descriptive statistics				
	کمینه Min	بیشینه Max	میانگین Mean	واریانس Variance	انحراف معیار Standard deviation
رس (Clay %)	8.00	53.60	31.08	70.96	8.42
سیلت (Silt %)	21.60	81.60	48.40	86.19	9.28
شن (Sand %)	3.20	65.60	20.53	164.09	12.81
شن خیلی ریز (Very fine sand %)	0.70	28.04	6.65	29.66	5.45
شن ریز (Fine sand %)	0.18	16.02	4.13	11.43	3.38
ماده‌ی آلی (Organic matter %)	0.22	6.33	2.58	1.67	1.29
کربنات کلسیم معادل (Calcium carbonate equivalent %)	1.00	80.65	25.32	445.90	21.12
شاخص پوشش گیاهی (NDVI)	-0.31	0.47	0.13	0.006	0.08
درصد شیب (Slope %)	3.10	80.61	28.31	273.88	15.42
جهت شیب (Aspect, radian)	8.87	365.85	171.82	10162.00	100.81
ارتفاع (Elevation, m)	907.00	2639.00	1714.00	88484.00	297.64

(شکل ۴) و نیز بر اساس مقادیر شاخص‌های آماری ارزیابی مدل مورد استفاده به نظر می‌رسد که مدل رگرسیونی خطی چند متغیره، فاقد دقت قابل قبول برای برآورد و مدل‌سازی پایداری خاکدانه‌ها در منطقه‌ی مورد مطالعه است. در واقع مدل‌های رگرسیون خطی تنها قادر به تشخیص روابط خطی بین متغیرهای مستقل و وابسته هستند و چنانچه روابط غیرخطی بین این متغیرها وجود داشته باشد، کارایی این مدل‌ها به حد قابل ملاحظه‌ای کاهش می‌یابد (۲۰ و ۲۸). بنابراین، به نظر می‌رسد که دلیل اصلی عملکرد ضعیف مدل رگرسیون خطی در برآورد ویژگی مورد مطالعه در این پژوهش، روابط غیرخطی بین این ویژگی و ورودی‌های مورد استفاده در مدل‌سازی باشد.

#### مدل ماشین‌های بردار پشتیبان (SVMs)

مقادیر  $r$ ، MSE و ERROR% برای مدل SVM طراحی شده برای برآورد GMD به ترتیب برابر ۰/۸۶، ۰/۰۰۵ و ۱۰/۶ درصد بودند (جدول ۲). همچنین، ضریب تبیین نسبتاً مناسبی ( $R^2=0/74$ ) بین مقادیر مشاهده‌ای و شبیه‌سازی شده برای داده‌های تست مدل SVM حاصل شد (شکل ۵)؛ اگرچه در بسیاری از نقاط، بیش‌برآوردی از GMD وجود داشت (شکل ۶).

به‌علاوه، آموزکتا (۱) اظهار می‌دارد که آهک به سبب اثرات سیمان‌کننده‌ی خود می‌تواند از تخریب خاکدانه‌ها جلوگیری کند. کانتون و همکاران (۷) بیان کردند که ویژگی‌های توپوگرافیکی می‌توانند به صورت غیرمستقیم و از طریق اثر بر کانی‌شناسی خاک، رطوبت خاک، کربن آلی، بافت خاک و استقرار پوشش گیاهی بر ساختمان خاک و پایداری خاکدانه‌ها اثرگذار باشند.

#### مدل رگرسیون خطی چند متغیره (MLR)

مقادیر شاخص‌های آماری مورد استفاده برای ارزیابی کارایی مدل رگرسیونی طراحی شده برای برآورد GMD در جدول ۲ آورده شده‌اند. مقادیر  $r$ ، ERROR% و MSE برای مدل رگرسیونی طراحی شده برای برآورد GMD به ترتیب برابر ۰/۳۲، ۱۵/۷ درصد و ۰/۰۱۲ بودند. مقادیر مشاهده‌ای و شبیه‌سازی شده‌ی GMD برای داده‌های تست مدل نیز در شکل ۳ مقایسه شده‌اند. با توجه به این شکل، مشخص است که در برخی از نقاط به‌صورت غیریکنواخت، بیش‌برآوردی از GMD صورت گرفته است و در برخی دیگر از نقاط، GMD کمتر از مقادیر مشاهده‌ای آن‌ها تخمین زده شده است. با توجه به ضریب تبیین پایین ( $R^2=0/10$ ) بین مقادیر مشاهده‌ای و شبیه‌سازی شده برای داده‌های تست مدل رگرسیونی طراحی شده در این پژوهش

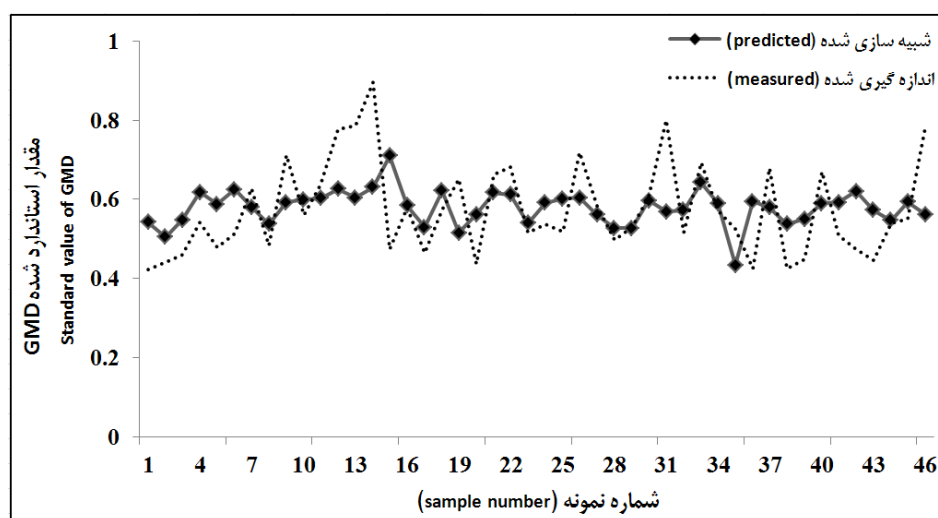
جدول ۲- شاخص‌های آماری مورد استفاده برای ارزیابی کارایی مدل‌های رگرسیون خطی چند متغیره (MLR) و ماشین‌های بردار پشتیبان (SVMs) در برآورد GMD

Table 2- Statistical criteria used to evaluate the performance of the MLR and SVM models for the prediction of GMD (Evaluation Criterion) شاخص آماری

نوع مدل (Model type)	ERROR%	MSE	$r$
MLR	15.70	0.012	0.32
SVMs	10.60	0.005	0.86

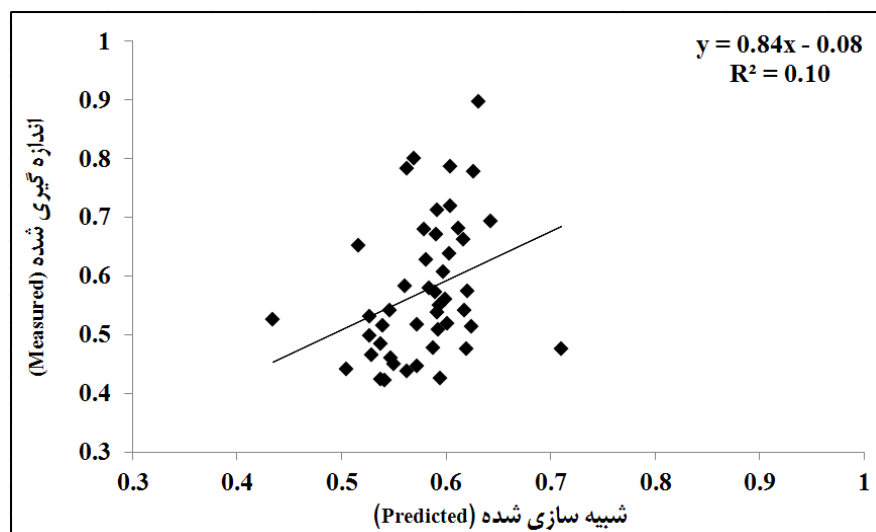
$r$ : ضریب همبستگی، MSE: میانگین مربعات خطا و ERROR%: درصد خطا

$r$ : correlation coefficient; MSE: mean square error; and ERROR%: error percentage



شکل ۳- مقایسه مقادیر مشاهده‌ای و شبیه‌سازی شده‌ی GMD برای داده‌های تست مدل رگرسیونی

Figure 3- Comparison of the observed and predicted GMD values for the test data of the regression model



شکل ۴- همبستگی بین مقادیر مشاهده‌ای و شبیه‌سازی شده‌ی GMD برای داده‌های تست مدل رگرسیونی

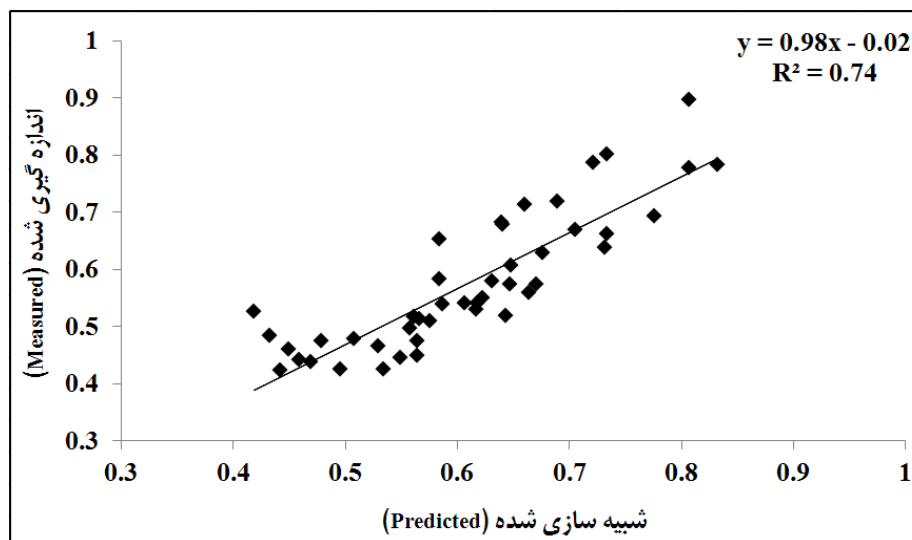
Figure 4- Correlation between the observed and predicted GMD values for the test data of the regression model

بیشتری نسبت به مدل رگرسیونی در برآورد و مدل‌سازی GMD در منطقه‌ی مورد مطالعه داشته باشند. در واقع، ماشین‌های بردار پشتیبان

بنابراین به نظر می‌رسد که بر اساس مقادیر نسبتاً قابل قبول شاخص‌های آماری ارزیابی مدل، ماشین‌های بردار پشتیبان دقت

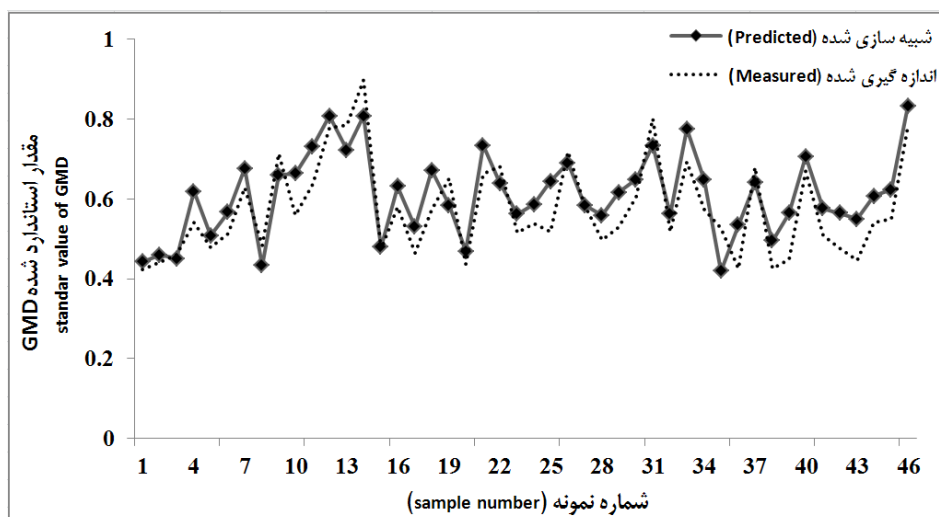
از دیگر عوامل مؤثر بر قابلیت مدل‌ها در برآورد و مدل‌سازی GMD در این پژوهش می‌باشند که به نظر می‌رسد مدل ماشین‌های بردار پشتیبان در مقایسه با مدل رگرسیون خطی دارای انعطاف‌پذیری بیشتری بوده است. بنابراین به نظر می‌رسد که در مناطقی که پیچیدگی‌های محیطی زیاد و بین داده‌های ورودی و خروجی، روابط غیرخطی پیچیده وجود داشته و درعین حال تعداد داده‌های قابل دسترس کم می‌باشد، روش‌های هوش مصنوعی مانند ماشین‌های بردار پشتیبان نسبت به روش‌های مرسوم مدل‌سازی مانند رگرسیون-های خطی برتری دارند.

ساخته شده با تابع کرنلی، دارای انعطاف‌پذیری زیاد و قابلیت بالایی در تشخیص روابط غیرخطی و پیچیده‌ی بین داده‌های ورودی و خروجی می‌باشند (۱۵). از سوی دیگر، روش‌های مبتنی بر آموزش در مقایسه با مدل‌های رگرسیون خطی از توانایی بیشتری در تشخیص روابط بین ورودی‌ها و خروجی برخوردار می‌باشند. همچنین روش‌های هوش مصنوعی مانند ماشین‌های بردار پشتیبان در مقایسه با روش‌های قدیمی و مرسوم مدل‌سازی مانند رگرسیون‌های خطی دارای قابلیت بالایی در تشخیص روابط پیچیده‌ی بین داده‌های ورودی و خروجی با تعداد نمونه‌ی کم هستند. پراکنش نمونه‌ها، تغییرات مکانی نقاط نمونه‌برداری و مقیاس منطقه‌ی مورد بررسی نیز



شکل ۵- همبستگی بین مقادیر مشاهده‌ای و شبیه‌سازی شده‌ی GMD برای داده‌های تست مدل SVM

Figure 5- Correlation between the observed and predicted GMD values for the test data of the SVM model



شکل ۶- مقایسه‌ی مقادیر مشاهده‌ای و شبیه‌سازی شده‌ی GMD برای داده‌های تست مدل SVM

Figure 6- Comparison of the observed and predicted GMD values for the test data of the SVM model



بردار پشتیبان دقت و قابلیت بیشتری نسبت به مدل‌های رگرسیونی مرسوم در برآورد و مدل‌سازی این ویژگی دارند. از سوی دیگر، با توجه به همبستگی نسبتاً مناسب بین مقادیر مشاهده‌ای و شبیه‌سازی شده‌ی GMD برای داده‌های تست مدل SVM طراحی شده به‌نظر می‌رسد که بتوان از ویژگی‌های زودیافت مورد استفاده در این پژوهش به‌عنوان ورودی SVM در مدل‌سازی و تخمین پایداری خاکدانه‌ها در منطقه‌ی مورد بررسی بهره‌جست. لذا این احتمال وجود خواهد داشت که بتوان از ماشین‌های بردار پشتیبان به‌عنوان روشی نوین برای برآورد ویژگی‌های فیزیکی و مکانیکی خاک برای رسیدن به بیشترین دقت با صرف کمترین هزینه و زمان استفاده نمود. با این وجود، بررسی امکان استفاده از تکنیک‌های انتخاب ورودی‌های بهینه نظیر الگوریتم‌های تکاملی<sup>۱</sup> (همچون الگوریتم ژنتیک<sup>۲</sup>) و نیز استفاده از دیگر ابزارهای مدل‌سازی در برآورد پایداری خاکدانه‌ها (مانند شبکه‌های عصبی مصنوعی و سیستم‌های استنتاج نروفازی انطباقی<sup>۳</sup>) در پژوهش‌های آینده پیشنهاد می‌گردد.

لامورسکی و همکاران (۱۳) از ماشین‌های بردار پشتیبان برای برآورد هدایت هیدرولیکی خاک استفاده نمودند و به قابلیت بالای ماشین‌های بردار پشتیبان در مدل‌سازی ویژگی‌های خاک اشاره کردند. ونگ و همکاران (۲۷) امکان استفاده از روش‌های گوناگون برآورد روان‌آب ماهیانه را بررسی و مقایسه نمودند و دریافتند که ماشین‌های بردار پشتیبان تخمین بهتری از روان‌آب ماهیانه (چه در مرحله‌ی آموزش و چه در مرحله‌ی اعتبارسنجی) در دسترس قرار دادند. تواراکاوی و همکاران (۲۲) نیز به توانایی بالاتر مدل‌های توسعه داده شده بر اساس ماشین‌های بردار پشتیبان برای برآورد پارامترهای هدایت هیدرولیکی در مقایسه با مدل‌های توسعه داده شده بر اساس شبکه‌های عصبی مصنوعی اشاره نمودند.

### نتیجه‌گیری کلی

نتایج این پژوهش، نمایانگر آن بود که مدل رگرسیون خطی چند متغیره، فاقد دقت قابل قبول برای برآورد و مدل‌سازی GMD در حوزه آبخیز کارون شمالی می‌باشد. این در حالی است که ماشین‌های

### منابع

- 1- Amezketta E. 1999. Soil aggregate stability: A review. *Journal of Sustainable Agriculture*, 14: 83-151.
- 2- Andesodun J.K., Mbagwu J.S.C., and Oti N. 2005. Distribution of carbon, nitrogen and phosphorus in water-stable aggregates of an organic waste amended Ultisol in southern Nigeria. *Bioresour Technology*, 96: 509-516.
- 3- Besalatpour A., Hajabbasi M.A., Ayoubi S., Afyuni M., Jalalian A., and Schulin R. 2012. Soil shear strength prediction using intelligent systems: artificial neural networks and adaptive neuro-fuzzy inference system. *Journal of Soil Science and Plant Nutrition*, 58: 149-160.
- 4- Besalatpour A.A., Ayoubi S., Hajabbasi M.A., Gharipour A., and Yousefian Jazi A. 2014. Feature selection using parallel genetic algorithm for the prediction of geometric mean diameter of soil aggregates by machine learning methods. *Arid Land Research and Management*, 28:383-394.
- 5- Calero N., Barron V., and Torrent J. 2008. Water dispersible clay in calcareous soils of southwestern Spain. *Catena*, 74: 22-30.
- 6- Canasveras J.C., Barron V., Del Campillo M.C., Torrent J., and Gomez J.A. 2010. Estimation of aggregate stability indices in Mediterranean soils by diffuse reflectance spectroscopy. *Geoderma*, 158: 78-84.
- 7- Canton Y., Sole-Benet A., Asensio C., Chamizo S., and Puigdefabregas J. 2009. Aggregate stability in range sandy loam soils: relationships with runoff and erosion. *Catena*, 77: 192-199.
- 8- Chenu C., Le Bissonnais Y., and Arrouays D. 2000. Organic matter influence on clay wettability and soil aggregate stability. *Soil Science Society of American Journal*, 64:1479-1486.
- 9- Dexter A.R., and Kroesbergen B. 1985. Methodology for determination of tensile strength of soil aggregates. *Journal of Agricultural Engineering Research*, 31: 139-147.
- 10- Gee G.W., and Bauder J.W. 1986. Particle size analysis. In: Klute, A. (Ed.), *Methods of Soil Analysis: Part 1., Agronomy Handbook No 9., American Society of Agronomy and Soil Science Society of America, Madison, WI*, pp. 383-411.
- 11- Kemper W.D., and Rosenau K. 1986. Size distribution of aggregates. In: Klute, A. (Ed.), *Methods of Soil Analysis: Part 1: Physical and Mineralogical Methods, American Society of Agronomy, Madison, WI*, pp. 425-442.
- 12- Khalilmoghadam B., Afyuni M., Abbaspour K.C., Jalalian A., Dehghani A.A., and Schulin R. 2009. Estimation of

- 1- Evolutionary algorithms
- 2- Genetic algorithm
- 3- Adaptive Neuro-Fuzzy Inference Systems, ANFIS

- surface shear strength in Zagros region of Iran-A comparison of artificial neural networks and multiple-linear regression models. *Geoderma*, 153: 29-36.
- 13- Lamorski K., Pachepsky Y., Slawinski C., and Walczak R.T. 2008. Using support vector machines to develop pedotransfer functions for water retention of soils in Poland. *Soil Science Society of American Journal*, 72: 1243-1247.
  - 14- Lamorski K., Pastuszka T., Krzyszcak J., Slawinski C., and Witkowska-Walczak B. 2013. Soil water dynamic modeling using the physical and support vector machine methods. *Vadose Zone Journal*, 12 (4):
  - 15- Li H., Liang Y., and Xu Q. 2009. Support vector machines and its applications in chemistry. *Chemometrics Intelligent Laboratory Systems*, 95: 188-198.
  - 16- Liao K., Xu S., Wu J., Zhu Q., and An L. 2014. Using support vector machines to predict cation exchange capacity of different soil horizons in Qingdao City, China. *Journal of Plant Nutrition and Soil Science*, 177 (5): 775-782.
  - 17- Nelson D.W. and Sommers L.E. 1982. Total carbon, organic carbon, and organic matter. In: Page, A. L. (Ed.), *Methods of Soil Analysis*, American Society of Agronomy, Madison, Wis, pp. 539-579.
  - 18- Nelson R.E. 1982. Carbonate and gypsum. In: Page, A.L. (Ed.), *Methods of Soil Analysis: Part I: Agronomy Handbook No 9*, American Society of Agronomy and Soil Science Society of America, Madison, WI, pp. 181-197.
  - 19- Six J., Paputian K., Elliot E.T., and Combrink C. 2001. Soil structure and organic matter. I. Distribution of aggregate-size classes and aggregate-associated carbon. *Soil Science Society of American Journal*, 64: 681-689.
  - 20- Sobhani J., Najimi M., Pourkhorshidi A.R., and Parhizkar T. 2010. Prediction of the compressive strength of no-slump concrete: A comparative study of regression, neural network and ANFIS models. *Journal of Construction and Building Materials*, 24: 709-718.
  - 21- Tisdall J. M., and Oades J.M. 1982. Organic matter and water-stable aggregates in soils, *Soil Sci.*, 33: 141-163.
  - 22- Twarakavi N.K.C., Simunek J., and Schaap M.G. 2009. Development of pedotransfer functions for estimation of soil hydraulic parameters using support vector machines. *Soil Science Society of American Journal*, 73: 1443-1452.
  - 23- Van Bavel C.H.M. 1950. Mean-weight diameter of soil aggregates as a statistical index of aggregation. *Proceeding of Soil Science Society of American*, 14: 20-23.
  - 24- Vapnik V. 1995. *The Nature of Statistical Learning Theory*, Springer-Verlag, New York.
  - 25- Vapnik V. 1998. *Statistical Learning Theory*, Wiley, New York.
  - 26- Wang L. 2005. *Support Vector Machines: Theory and Applications*. Springer-Verlag, New York.
  - 27- Wang W.C., Chau K.W, Cheng C.T., and Qiu L. 2009. A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. *Journal of Hydrology*, 374: 294-306.
  - 28- Yilmaz I., and Yuksek G. 2009. Prediction of the strength and elasticity modulus of gypsum using multiple regression, ANN, and ANFIS models. *International Journal of Rock Mechanics and Mining Sciences*, 46: 803-810.



## Modeling of Soil Aggregate Stability using Support Vector Machines and Multiple Linear Regression

A.A. Besalatpour<sup>1\*</sup> - H. Shirani<sup>2</sup> - I. Esfandiarpour Borujeni<sup>3</sup>

Received: 09-06-2013

Accepted: 05-04-2015

**Introduction:** Soil aggregate stability is a key factor in soil resistivity to mechanical stresses, including the impacts of rainfall and surface runoff, and thus to water erosion (Canasveras et al., 2010). Various indicators have been proposed to characterize and quantify soil aggregate stability, for example percentage of water-stable aggregates (WSA), mean weight diameter (MWD), geometric mean diameter (GMD) of aggregates, and water-dispersible clay (WDC) content (Calero et al., 2008). Unfortunately, the experimental methods available to determine these indicators are laborious, time-consuming and difficult to standardize (Canasveras et al., 2010). Therefore, it would be advantageous if aggregate stability could be predicted indirectly from more easily available data (Besalatpour et al., 2014). The main objective of this study is to investigate the potential use of support vector machines (SVMs) method for estimating soil aggregate stability (as quantified by GMD) as compared to multiple linear regression approach.

**Materials and Methods:** The study area was part of the Bazoft watershed (31° 37' to 32° 39' N and 49° 34' to 50° 32' E), which is located in the Northern part of the Karun river basin in central Iran. A total of 160 soil samples were collected from the top 5 cm of soil surface. Some easily available characteristics including topographic, vegetation, and soil properties were used as inputs. Soil organic matter (SOM) content was determined by the Walkley-Black method (Nelson & Sommers, 1986). Particle size distribution in the soil samples (clay, silt, sand, fine sand, and very fine sand) were measured using the procedure described by Gee & Bauder (1986) and calcium carbonate equivalent (CCE) content was determined by the back-titration method (Nelson, 1982). The modified Kemper & Rosenau (1986) method was used to determine wet-aggregate stability (GMD). The topographic attributes of elevation, slope, and aspect were characterized using a 20-m by 20-m digital elevation model (DEM). The data set was divided into two subsets of training and testing. The training subset was randomly chosen from 70% of the total set of the data and the remaining samples (30% of the data) were used as the testing set. The correlation coefficient ( $r$ ), mean square error (MSE), and error percentage (ERROR%) between the measured and the predicted GMD values were used to evaluate the performance of the models.

**Results and Discussion:** The description statistics showed that there was little variability in the sample distributions of the variables used in this study to develop the GMD prediction models, indicating that their values were all normally distributed. The constructed SVM model had better performance in predicting GMD compared to the traditional multiple linear regression model. The obtained MSE and  $r$  values for the developed SVM model for soil aggregate stability prediction were 0.005 and 0.86, respectively. The obtained ERROR% value for soil aggregate stability prediction using the SVM model was 10.7% while it was 15.7% for the regression model. The scatter plot figures also showed that the SVM model was more accurate in GMD estimation than the MLR model, since the predicted GMD values were closer in agreement with the measured values for most of the samples. The worse performance of the MLR model might be due to the larger amount of data that is required for developing a sustainable regression model compared to intelligent systems. Furthermore, only the linear effects of the predictors on the dependent variable can be extracted by linear models while in many cases the effects may not be linear in nature. Meanwhile, the SVM model is suitable for modelling nonlinear relationships and its major advantage is that the method can be developed without knowing the exact form of the analytical function on which the model should be built. All these indicate that the SVM approach would be a better choice for predicting soil aggregate stability.

**Conclusion:** The pixel-scale soil aggregate stability predicted that using the developed SVM and MLR models demonstrates the usefulness of incorporating topographic and vegetation information along with the soil properties as predictors. However, the SVM model achieved more accuracy in predicting soil aggregate stability compared to the MLR model. Therefore, it appears that support vector machines can be used for prediction of

1, 2, 3- Assistant Professor, Associate Professor and Assistant Professor, Department of Soil Science, Collage of Agriculture, Vali-e-Asr University of Rafsanjan, Respectively  
(\*-Corresponding Author Email: a.besalatpour@vru.ac.ir)

some soil physical properties such as geometric mean diameter of soil aggregates in the study area. Furthermore, despite the high predictive accuracy of the SVM method compared to the MLR technique which was confirmed by the obtained results in the current study, the advantages of the SVM method such as its intrinsic effectiveness with respect to traditional prediction methods, less effort in setting up the control parameters for architecture design, the possibility of solving the learning problem according to constrained quadratic programming methods, etc., should motivate soil scientists to work on it further in the future.

**Keywords:** Geometric mean diameter (GMD), Soil structure, Support vector machines (SVMs), Soil physical properties